DOCUMENT RESUME

ED 094 481 EA 006 327

AUTHOR Theimer, William C., Jr.
TITLE Problems in the Evaluation of IMPACT on Experimental Schools.
PUB DATE Apr 74
NOTE 13p.; Paper presented at American Educational Research Association Annual Meeting (59th, Chicago, Illinois, April 15-19, 1974)

EDRS PRICE MF-$0.75 HC-$1.50 PLUS POSTAGE
DESCRIPTORS *Educational Research; Elementary Schools; Evaluation; *Evaluation Methods; *Evaluation Needs; *Experimental Schools; Problems; *Program Evaluation; Research Utilization; Secondary Schools
IDENTIFIERS Experimental Schools Program

ABSTRACT

The National Institute of Education (NIE) funded 18 school districts to implement programs that were to result in 'holistic' change. Two kinds of data were expected --those generated to measure the impact of the Experimental Schools Program and those generated by field studies, which were discussed in other symposia. In measuring impact by the level II contractors, two major problems faced researchers: the omission, in many instances, of qualified educational researchers in the selection of personnel to staff the various projects; conflict, over-testing of local staff and students, and wasted effort on the part of both teams of evaluators, which resulted from the forced separation of summative and formative evaluation between two different groups; and some lack of required data and the acquisition of useless data, which resulted from a failure to state clearly the objectives in measurable terms. (Author/EA)

ED 094481

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PROBLEMS in the EVALUATION of IMPACT of EXPERIMENTAL SCHOOLS

Speech presented at AERA Convention
Chicago - April 16, 1974
by

Wm. C. Theimer, Jr.
University of Southern California

BEST COPY AVAILABLE

Let me begin with a little background on the Experimental Schools Program. First, Congress has been besieged in recent years by many people who have suggested that major changes needed to be made in the education process in America (Holt, 1964; Hentoff, 1967; & Silberman, 1970). It has also been evident that major changes in schooling practices in America have not occurred (Sizer, 1973). One of the reasons given for the lack of change has been that the efforts exerted to effect changes have been done on a 'piecemeal' basis rather than as a total review and restructuring of a school system. Such a comprehensive change would include community involvement in determining what the goals of its schools were, how congruent these goals were with existing practice in the classroom, and planning to effect congruence where it was not evident. Students should be involved also to insure that the education they were receiving was 'relevant' and meeting both their immediate and long-range needs. Congress has also heard from educational researchers who have consistently pleaded that more monies were needed for research, evaluation and development purposes, and that any attempt to restructure education should include both monies and plans for adequate evaluation and research into the change process itself and the effects of the change (Hayman, 1968; Mouly, 1971; & Borg, 1973). It has also been recognized by these researchers that more substantive efforts were needed to improve the methodologies and techniques currently used in educational evaluations.

In an effort to meet these criticisms and needs, the Experimental Schools Program (ESP) was devised within the U.S.O.E. It was designed to provide

EA 006 327

school districts with enough funds to effect major changes in their school systems and at the same time to provide enough monies to thoroughly research and evaluate both the process of change and the effects of that change. In answer to the criticism that most educational research had been 'piecemeal', funding was to be generous and guaranteed for five years; and to answer the criticism regarding research monies, twenty-five per cent (25%) of the total funds were to be allocated for research and evaluation purposes.

Given this plan for producing comprehensive change in the system, a concept of evaluation involving three levels was also introduced. The Level I team was to be a part of the local ESP staff, and was to perform those functions of evaluation which would be required of any local school research office; thus their function was to be largely formative. They were to provide a cybernetic type of feedback which would help the project to achieve its goals by gathering, analyzing, and presenting data regularly to the local school staff in terms of the effectiveness of the new programs being implemented. They were also to be responsible for the regular testing program, and the presentation of test data annually to the local Boards, staff, and other interested parties.

The Level II team was to be funded directly by NIE, and was to be a team of 'outside evaluators' working closely with the project, but not part of the local staff. They were to be located on-site, however. Those firms who received the contracts to implement the Level II concept were to provide the NIE with regular reports on the progress of the project, including initially reports on the Level I team effectiveness. This latter function was dropped soon after the projects began for reasons which will become evident later. These reports by the Level II team were to be summative in nature, and to measure the effects of 'holistic' change. To insure that as much of the change as possible could be captured, the teams were to be

interdisciplinary, and to include such professionals as anthropologists or sociologists or political scientists or economists. Through this mix of disciplines, it was believed that new methodologies and techniques for evaluating education would emerge.

The initial plans called for Level III to be composed of a group of highly esteemed Educational Researchers who would function in a role much like that of Independent Educational Auditors. One plan was to have a group such as AERA sponsor this auditing group, but for various reasons, these plans did not materialize. At present the NIE consultants are acting in this auditing function, and providing the guidance considered necessary to accomplish the goals of the ESP.

With these plans developed, proposals were requested from school districts across the nation to implement changes which were felt to be needed within the system. Eighteen were accepted and funded. Five of these were in suburban areas, including the two sites represented on the stage here - San Antonio, Texas and Greer, South Carolina. The others are located in Berkeley, California; Minneapolis, Minnesota; and Tacoma, Washington. Three urban sites, under the direction of the Urban League, and ten rural sites were also funded. Once the sites were selected, contractors were asked to bid to carry out the Level II evaluation once the sites had been determined, with the rural sites being evaluated as one unit by one contractor. Although most of my information has come from the five suburban sites, much of what will be said will apply to the other sites as well.

At each site of the five mentioned the system to be included in the ESP was composed of from four to six elementary schools, their related middle school and high school. The plans submitted by the local districts called for substantial change in educational practices, and were aimed at goals such as improvement of education of integrated schools, individualization

of education, and creating an education compatible with the local culture. Basic to all of these, however, was the goal to effect 'holistic' change in the system.

One other distinctive feature of the ESP should be made before the problems involved in evaluating Impact are discussed. Included in the scheme for evaluating ESP were two different approaches - Field Studies and Impact Studies. Field Studies were to be descriptive and basically anthropological in nature, and were included to try to get at some of the more intangible kinds of effects the schools have which humanistic psychologists feel are not measured in traditional kinds of evaluations. Impact studies, on the other hand, were to determine the impact of the project on all the participants and groups involved in the ESP. This paper will confine itself to those problems involved in the measurement of Impact. Other symposia and papers will address themselves to the problems encountered in developing Field Studies evaluations.

The word Impact has a fine ring to it, but when efforts were made to define the term operationally, the complaint was raised that the resulting evaluation was pedestrian and mediocre, and that it missed the essence of what the project was all about. This complaint is a common one (Combs, 1973), but efforts to measure the intangibles which are hard to define operationally have proved to be not too successful to date; therefore, one of the major goals of the ESP was to devise new, more comprehensive kinds of evaluation designs and techniques. The original evaluation schemes submitted to NIE by Level II staffs proved to be either typical of what has been done in the past or so grandiose and immeasurable that they were impractical. It finally became apparent that if some baseline data was to be obtained the first year of the project that some kind of validated measures had to be obtained, even if they were pedestrian, and that the

development of more esoteric evaluation schemas would have to be devised as the projects began to function. This decision resulted in the use of traditional kinds of evaluation designs, based on the CIPP model of Stufflebeam (1968), the Discrepancy Model of Provus (1968), and the Judgmental Model of Stake (1967). The measures taken were from such traditional assessment devices as standardized achievement tests and the OCDQ.

Given this background, what are the major problems facing the evaluators in measuring Impact at all levels? First, the idea of separating the formative from the summative evaluation in Level I and Level II respectively has not proved feasible. The reports generated by Level II will, in spite of any attempts to channel them only to NIE or any other group, have an impact on the Project. The mere presence of the Level II team on the scene will have some effect on the project, and it has become the practice that Level II reports are shared with the Level I team, which then has the responsibility of disseminating the information thus obtained to project personnel as they see fit. Since both groups frequently need the same or similar data, and must obtain their data from the same groups, it is important that they work together. Thus, the artificial separation with the Level II team operating as policemen simply proved unworkable. In practice the groups are working together to develop questionnaires, interview schedules, and to plan sampling strategies to a greater or lesser extent at all sites. Although this problem has been partially resolved, there are still areas in which the Level II team has difficulty obtaining needed information because the Level I team does not feel that the information should be provided to them.

However, since they are required to obtain data from the same groups, and since both have manpower shortages in terms of the kinds of information required of them, it is essential that they work together. In an even

more important sense, they must work together since they can pool their talents in the development of questionnaires, interview schedules, and other evaluative measures and techniques. By pooling their resources, better data gathering instruments can be developed; and even more important, the groups from which the data is obtained can be spared the onus of being tested or interviewed by two different groups for information which is largely redundant. Although this problem is being resolved, it cannot reach complete resolution until the second problem has been solved.

The second problem involves the instability of staffs at Level I and Level II sites, and within NIE. The turnover of Level II staff at most of the sites has been over 50%, and has reached 100% in some. The problems this instability has created can be readily understood. Level I staffing has also been plagued with turnovers, and in some instances, no staffing at all. These difficulties have been compounded when project personnel at NIE change, and these changes are reflected in changes in direction or emphasis. The problem is similar to that described by Dershimer in 1972 when he said that the Educational Labs were "plagued by shifting agency directives and requirements from change in leadership in OE and by constant auditing and administrative interference". Many of the difficulties caused by these changes in personnel at NIE can be overcome by having all directives made in written form, and I believe that this policy has in fact been implemented recently.

The problem of excessive turnover of Level II staff has been caused by poor initial selection in many cases. That is, the original RFP emphasized the need for multi-disciplinary personnel and largely ignored the need for personnel trained in educational research. This emphasis led many of the contractors to include political scientists or economists or

others who had had either little or no experience in conducting research in the schools as directors of their Level II teams, and to exclude any person who was an educational researcher.

Although many people have either stated or inferred that educational research is not a fully developed discipline, it is fortunately true that we do have a discipline which has developed in recent years a set of strategies and techniques which work, albeit not as well as we would like. Too many people have had too high expectations of education in the past, as Chase pointed out in the Phi Delta Kappan in 1970, and we have suffered as a consequence (Sizer, 1973). Today, educational researchers are much less likely to discuss what they will find before they complete their evaluations than they were a few years ago, primarily because they are better trained. The influx of Cooperative Research and Title IV monies following Sputnik has up-graded the competencies of many educational researchers and created a pool of well-trained personnel at the same time. This training program, and the introduction of new models for evaluation in recent years, have provided a fairly substantial base from which better and more complete models can be developed to measure the more difficult kinds of learnings and behaviors which we do not adequately assess now. It is still unfortunately true, however, that many people 'hire out' as educational researchers or evaluators with minimal or no training in the discipline. The reasons for this are diverse, but probably are a result of the thinking that everyone is an expert in education. The problem is not unique to ESP, however. Shutz (1973) commented, "Concern for training and retraining of research personnel in education has historically lagged federal initiatives in educational R & D programs. This lag had had vicious consequences.

Planning for training has been too little and too late, and support for training, when provided, has been too much and too soon. The cycle has gone like this: "A man-power demand is created as new programs are launched. By the time the demand has been noted, the programs are already in difficulty. Training programs are then created. By the time they are operative, the demand is absent, for the R & D initiatives have been abandoned as failures."

This pattern seemed to be operating when the original contractors were funded since some of the original staffs did not include even one educational researcher. While no one will object to the need for a multi-disciplinary effort, the omission of educational research personnel on these teams was short-sighted and resulted in inadequate evaluative designs since those developing them were not acquainted with the kinds of problems and constraints which had led to the development of the models currently used widely in education. The problem was one of starting from scratch instead of building on the work of others who had been faced with similar tasks; of conflict with local school authorities; and of inaction. These difficulties have resulted in the selection of personnel on the Level II teams which are presently operating who are knowledgeable and trained in educational research, generally. There is still a need for more highly trained personnel on some staffs to adequately measure Impact, but the present staffs are performing well. The use of consultants who are experienced has enabled some personnel to learn 'on the job,' and provided some insight into the trouble spots to be expected on the local scene so that the problems could be avoided. It would appear to be desireable, however, that at least one member of the staff on any Level II team be well-versed in educational research simply to avoid the kinds of duplication of effort which occurs when a person experienced in one field steps into another in

which he is not up-to-date. The point I wish to make here is that not any researcher is qualified to be an educational researcher. We do possess unique kinds of knowledge and methodology which differentiate us from other researchers.

A second problem which this pattern of staff selection and instability has created is that of trying to do too much too soon with too few competent people. One of the goals of NIE in funding the ESP's was to develop new methodologies, new research techniques, and new measurement instruments through the activities of both Level I and Level II staff. Because of the problems involved in developing adequate evaluation schemes, little time has been left for either developing good data management and data gathering plans, or for the development of carefully planned and built measurement instruments or evaluation models. It is true that the Field Studies will develop a model for describing schools since they were deliberately designed to incorporate anthropological techniques and methods into school settings. The same is not true for Impact Studies, however, and the basic models mentioned earlier have largely been used in the development of evaluation schemes for collecting baseline data. This procedure seems to me to be a reasonable one since it utilizes existing knowledge as a base from which to develop new knowledge. One fact needs clarification, however. It is well known that the developers of standardized tests spend several years constructing and norming their instruments; the amount of staff time devoted to writing items, checking reliability, validity and format is considerable. That amount of time simply does not exist for Level II staff at present since they have spent much of their time writing papers and reports, and much of their time of necessity must be spent in observation if they are to monitor and document the progress of the ESPs. If the writing

required by NIE continues at the present level, very little new material or methodology will come from the ESP; if the level of writing decreases substantially now that the teams are functioning, then some new evaluation material, and techniques can be expected to come from this effort, such as testing material similar to that which Mr. Cervantes will describe or analytic processes such as Dr. Culver will present.

A final problem involves the measurement of Impact when there are no well-stated objectives in the proposed plan of action submitted by the local district. Although Scriven has discussed Goal-Free Evaluation techniques, he had not indicated that total evaluations be made on a goal free basis. Neither would it appear to be feasible to try to measure Impact on a goal free basis. It is true that the Field Studies, being largely descriptive, resemble the goal free type of evaluation Scriven has described. It would appear that Impact Studies must address themselves to the kinds of changes which the schools propose to make; that is, the impact of the ESP on the local community can best be assessed in terms of how well the project accomplishes its objectives. The problem has been that the objectives have been rather ambiguously stated in the initial proposals, being generally in the form of goals. This fact has resulted in the Level II team facing a dilemma - either they must define the objectives or wait until the local staff does the defining. If they pursue the former course, they may be accused of building 'straw men,' which they can then either build up or knock down. If they pursue the latter course, they may beinactive and end up with no data upon which to base any conclusions regarding the effectiveness of ESP at the end of the five year period. It is possible for Level II to define the goals objectively, then obtain agreement from the project staff and/or the community that their

interpretation was correct, but this process is slow and rarely leads to a consensus. The first year's evaluation designs suffered from this malady, and such terms as 'compatibility' are still not clearly defined, yet remain major goals of the projects. Level II personnel have reacted to the problem by writing position papers and developing evaluation strategies to assess the objectives described in the papers, while at the same time strongly urging the project staff and community to either concur with their position or interpret the goals in measureable terms themselves. Until the interpretation made by the Level II team is accepted by the local project staff and the community, the problem of whether or not the final conclusions drawn by Level II based on their data will reflect the precise goals of the project personnel and the community remains in doubt. Although some objectives have been defined by project personnel simply by what they have done over the course of the first year and a half, it is still true that some areas are not being adequately assessed because no well defined objectives have been obtained.

In conclusion, it appears that the problems associated with measuring the Impact of ESP have been partially resolved during this past year. There are still some major areas which need attention, but it appears that at least some effort is being expended to address them. There are some points that should be noted now, however. One, very little in the way of new techniques or methodologies have been developed to date. Two, much of the effort expended to date has been done by professionals other than educational researchers. Three, some projects have not had continuity in the development of evaluation plans, and the result will be that the final evaluation will consist of data covering a period of time less than five years. Four, it has taken approximately a year to a year and a half for the

projects to begin to function, the different Levels of evaluation to operate effectively, and the NIE to stabilize. What do these statements have to do with AERA?

Simply that these projects are the first to have major allocations made by Congress for research and evaluation purposes. As a result, the effectiveness of the evaluation effort will reflect on the educational research community whether we like it or not unless we clearly explicate the degree of our involvement now. Just as the energy crisis of last fall had people pointing their fingers at scientists as well as the government and the oil industry, so we will be held accountable for the results of the evaluation of the ESP. It is essential, it seems to me, that we be aware of the developments in the ESP as a major effort to improve educational R & D by NIE, and to emphasize the fact that failure of ESP to produce major improvements in educational research should not be used as evidence that monies spent for educational research are wasted, since we have been only minimally involved in the program to date.

REFERENCES

Borg, W. F. Educational Research. 2d ed. (New York, N. Y.: David McKay, 1971)

Combs, A. W. "Educational Accountability from a Humanistic Perspective". Educational Researcher, Sept. 1973, p. 19-21

Dershimer, R. A. "The Federal Commitment to R & D Institutions in Education". Educational Researcher, Dec. 1972, p. 3-4

Hayman, J. L. Jr. Research in Education. (Columbus, Ohio: Merrill, 1968)

Mouly, G. J. The Science of Education. (New York: Holt, Rinehart & Winston, 1971)

Scriven, M. "The Methodology of Evaluation". In B. O. Smith, ed. Perspectives of Curriculum Evaluation. (Chicago, Ill.: Rand, McNally 1967)

Shutz, R. E. "Now is the Future for R & D in Education". Educational Researcher, May 1973, p. 3

Sizer, T. R. Places for Learning, Places for Joy. (Cambridge, Mass.: Harvard University Press, 1973)

Stake, R. E. "The Countenance of Educational Evaluation". Teachers College Record. 68 (1967), pp. 523-540

Stufflebeam, D. I., Provus, M. et. al. Educational Evaluation and Decision Making. (Itasca, Ill.: F. E. Peacock Press, 1968)